THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 5, May–June 1981
Printed in U.S.A.

Adaptive Post-Filtering of ADPCM Speech

By N. S. JAYANT

(Manuscript received December 3, 1980)

We show that the quality of adaptive differential PCM *(*ADPCM*) speech can be significantly improved by passing it through a reconstruction low-pass filter that is matched to an appropriately defined short-time speech cutoff frequency. Practically, the adaptive procedure involves switching the decoder output into one of a bank of N low-pass filters whose cutoff frequencies span the expected range of input speech bandwidth. For the case of equally spaced filter cutoffs, and with uniform probability density function models for the quantization noise spectrum and the cutoff frequency, more than one-half of the maximum adaptive filtering gain is realizable by a bank of four filters. Computer simulations of 16- and 24-kilobit/s* ADPCM *coders indicate that perceived quality gains are in fact greater than what is indicated by an analytically predicted objective gain of 2.6 dB.*

I. SHORT-TIME CUTOFF FREQUENCY

ADPCM (adaptive-quantization/differential PCM) speech coding usually assumes a time-invariant model for speech bandwidth (such as 3200 Hz for telephone quality applications), and a corresponding time-invariant low-pass filter with cutoff $f_0 = 3200$ Hz for the decoded speech. However, short-time speech spectra of 3200-Hz-filtered speech exhibit cutoff frequencies f_c, that are often significantly smaller than the long-time nominal cutoff frequency f_0.[1] Figure 1 sketches a short-time spectrum at time t, and defines $f_c(t) = f_L^T(t)$ as the low-pass cutoff frequency that includes all but T percent of short-time spectral energy. Figure 2 shows long-time-averaged spectra for four sentence-length 3200 Hz band-limited utterances [L, B, C, and D], denoting ["A *l*athe is a big tool" (female utterance), "An icy wind raked the *b*each" (female utterance), "The *c*hairman cast three votes" (female utterance), and "This is a computer test of a *d*igital speech coder" (male utterance)], respectively. It is clear that all the four spectra are low

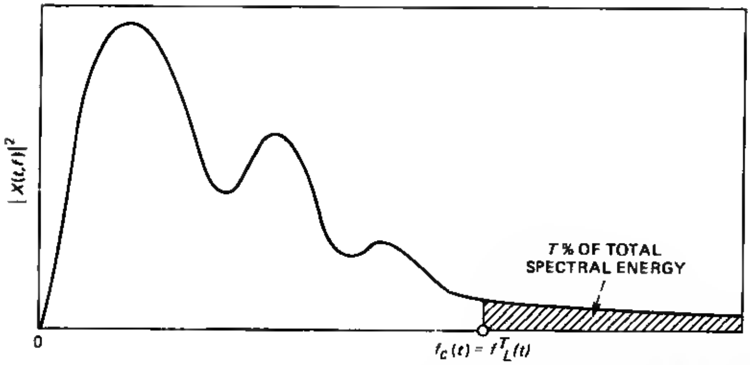

Fig. 1—Definition of short-time cutoff frequency $f^T_L(t)$. The shaded area includes T percent of short-time speech power.

pass in a long-time-average sense, a fact that is well exploited in fixed-prediction differential coding.[2] Figure 3 shows corresponding histograms of short-time cutoff frequency $f^T_L(t)$ for a threshold $T = 1$ percent. It is seen that on a short-time basis, speech segments can be either low pass (say, $f_L(t) < f_0/2$) or all pass (say, $f_L(t) > f_0/2$), although both of these segment types come from inputs that are low pass from a long-time-averaged energy viewpoint. It is also clear that the four histograms of Figure 3 are very different; however, as a single descriptor of these histograms, we propose a uniform probability density model for $f_L(t)$,

$$p(f^T_L(t)) = \frac{1}{f_0}, \qquad 0 < f_L(t) < f_0. \tag{1}$$

The adequacy of the above model is clearly a function of the threshold T. Clearly for the extreme cases of $T = 100$ percent and $T = 0$ percent, the pdf of $f^T_L(t)$ would degenerate into delta functions at $f = 0$ and f_0, respectively, with corresponding low-pass counts of 100 and 0 percent. The uniform density model (1), on the other hand, implies equal, 50 percent occurrences of low-pass and all-pass segments, and Fig. 4 shows that this is reasonable as a sentence-ensemble average for thresholds $T = 1$ and 2 percent. A threshold of $T = 1$ percent has been used in all of our ADPCM simulations.

The value of $T = 1$ percent produces a filtering distortion of $10 \log(1/100) = -20$ dB. More relevantly, it constitutes a "perceptually acceptable" low-pass threshold, with a filtering distortion that is obvious only in critical listening. On the other hand, at $T = 2$ percent, the filtering distortion begins to get obvious, and undesirable even for low-quality application such as 16 kilobit/s. The role of threshold T is discussed at greater length in a recent work which relates to the

refinement of ADM (adaptive delta modulation) with adaptive post-filtering.[1] This work also shows that consideration of the low-pass threshold $f_L^T(t)$ is more useful, for purposes of reducing coder noise, than the consideration of a high-pass threshold $f_H^T(t)$; in other words, providing a band-pass reconstruction filter matched to short-time high-pass and low-pass cutoffs $[f_H(t),\ f_L(t)]$ was not significantly better than providing a low-pass reconstruction filter matched to the range $[0,\ f_L(t)]$.

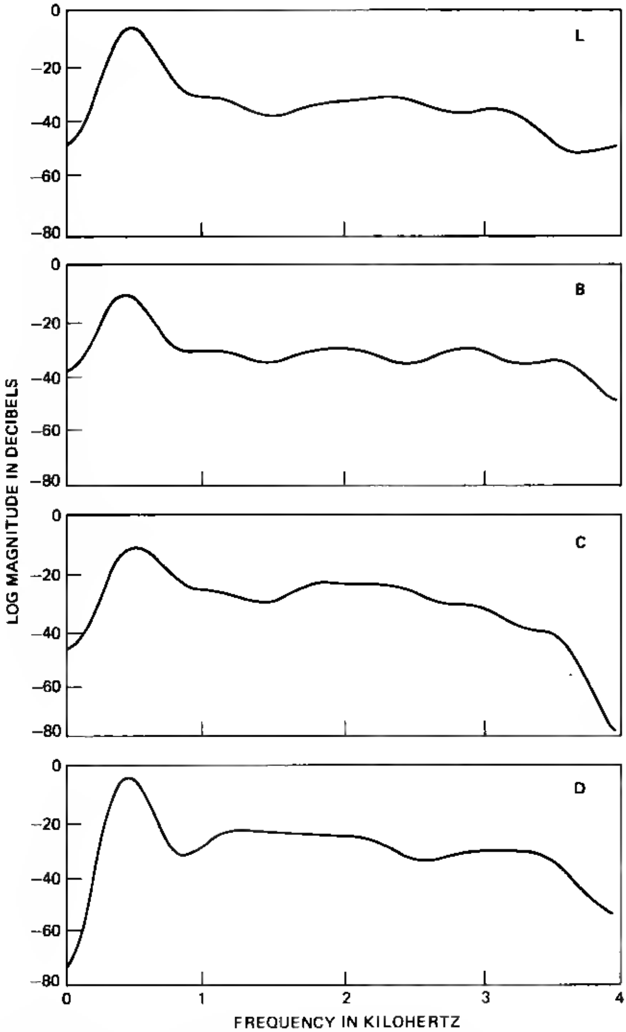

Fig. 2—Long-time averaged spectra for four sentence-length inputs. A 3.2-kHz band limitation is obvious in all four examples.

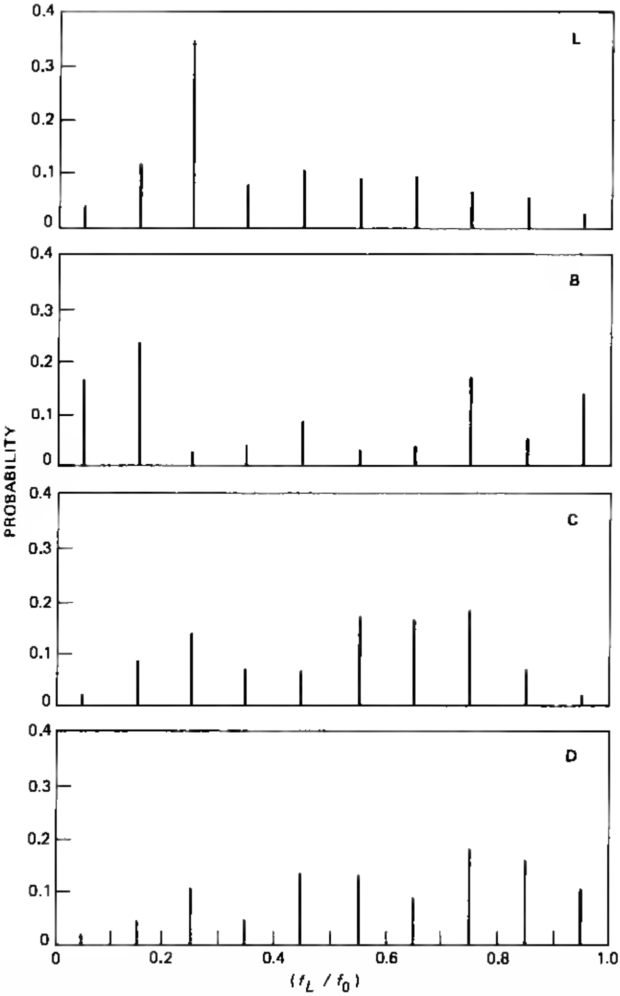

Fig. 3—Histograms of short-time cut-off frequency $f_L^T(t)$ ($T = 1$ percent) for the four inputs of Fig. 2. The short-time cutoff is in general significantly less than 3.2 kHz, the cutoff in the long-time-averaged spectra of Fig. 2.

II. ADAPTIVE LOW-PASS FILTERS

Figure 5a shows the effect of a low-pass reconstruction filter in ideal adaptive post-filtering; out-of-band ADPCM quantization noise components in the cross-hatched range $[f_L(t),\ f_0]$ in the noise spectrum are rejected by an ideal low-pass filter matched to $f_L(t)$.

Figures 5b and 5c show suboptimal but practical versions of 5a, where the value of $f_L(t)$ causes switching the decoder output into one of a bank of N low-pass filters, leading to a degree of noise-rejection

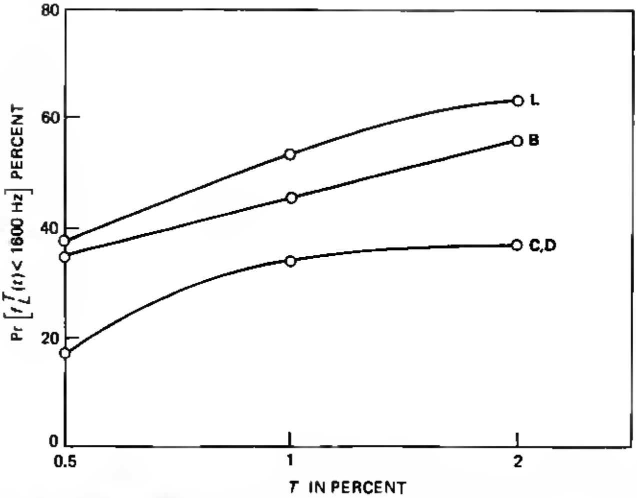

Fig. 4—Percentages of segments with $f_L^T(t) < f_0/2$ in the four inputs of Fig. 2. For $T = 1$ and 2 percent, the mean percentage of such segments is in the order of 50 percent.

that is always less than that in Fig. 5a. For example, in the $N = 2$ example of Fig. 5b, the filter bank consists of two filters with fixed cutoffs $f_{c2}(= f_0)$ and f_{c1}; in the upper illustration in Fig. 5b, the value of $f_L(t)$ is not small enough to switch in the lower filter f_{c1}; consequently there is no out-of-band noise rejection similar to that in the upper example of Fig. 5a. With the uniform pdf model (1), the two-filter bank system realizes out-of-band noise rejection only 50 percent of the time (when $f_L(t) < f_{c1} = f_0/2$). The four-filter system of Fig. 5c is clearly more effective; it realizes nonzero noise rejections in three out of the four cases shown, and indeed for 75 percent of all speech segments, if the uniform pdf (1) is valid.

A block diagram of an N-filter-bank adaptive system appears in Fig. 6. Note that for simplicity, the cutoff frequencies are equally spaced,

$$f_{cn} = f_0 \frac{n}{N}, \tag{2}$$

and that filter n is switched on when the input frequency cutoff is in the appropriate (f_0/N)-wide range,

$$\text{Switch to filter } n \quad \text{if} \quad \frac{n-1}{N} < \frac{f_L(t)}{f_0} \leq \frac{n}{N}. \tag{3}$$

III. SEGMENTAL s/n GAINS $G(N)$

The spectrum of quantization noise depends on many factors, including the nature of adaptive quantization, the input spectrum, and

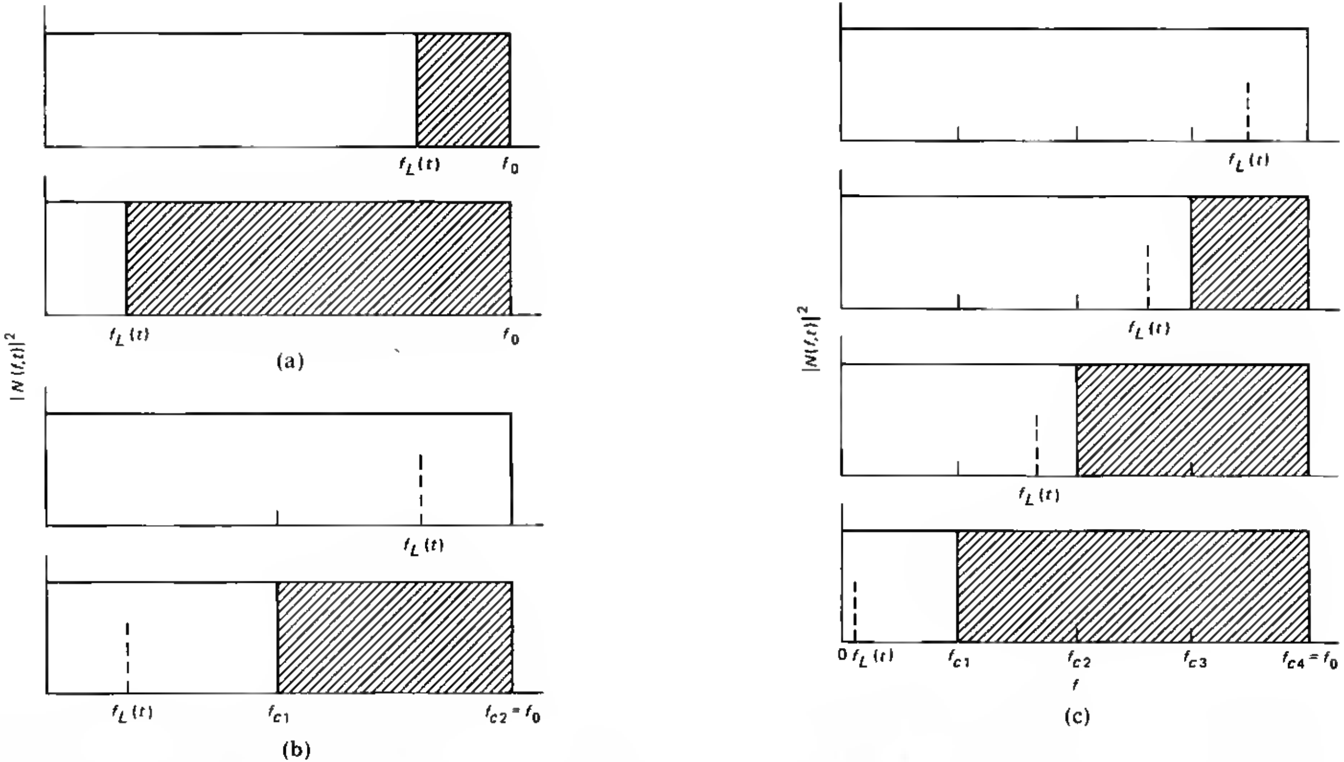

Fig. 5—Out-of-band noise rejection in (a) ideal adaptive filtering and in practical filter-bank schemes with (b) $N = 2$, and (c) $N = 4$ low-pass filters. Cross-hatched regions refer to rejected portions of noise spectrum. In the practical schemes (b) and (c), if f_{CL} refers to the allowed filter cutoff f_{cr} $(r = 1, 2, 3, 4)$ that is nearest to and higher than $f_L(t)$, the out-of-band noise in the frequency range $(f_L(t),\ f_{CL})$ is left unrejected.

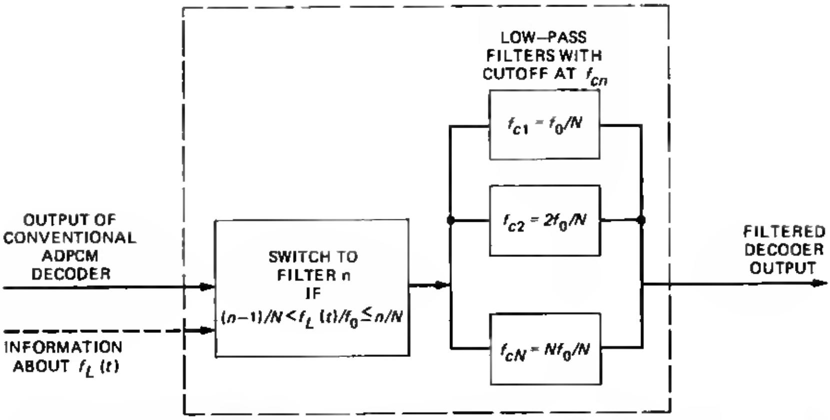

Fig. 6—Block diagram of adaptive low-pass filtering system with an N-filter bank. Allowed cutoff frequencies are equally spaced, being integral multiples of f_0/N. The entire dashed box constitutes an optional refinement of a conventional ADPCM decoder.

the effect of predictor; but experience indicates that the white-noise spectrum of Fig. 5 is indeed the single most reasonable model.[1,3] Combining this assumption with the uniform pdf for the cutoff frequency $f_L(t)$, we shall now develop expressions for expected gains in segmental[2] s/n due to adaptive post-filtering.

When the input cutoff $f_L(t)$ is such as to switch filter n ($n = 1, 2, \cdots, N$), the noise rejection factor is N/n (see Fig. 5), with a maximum value of N for the extremely low values of $f_L(t)$, and a minimum value of 1 for extremely high values of $f_L(t)$. The expected gain (in dB) is therefore given by

$$G(N) = \sum_{n=1}^{N} \left(10 \log \frac{N}{n}\right) \cdot \left(\Pr[n] = \frac{1}{N}\right) \text{dB}, \tag{4}$$

where the probability of switching in filter n is $\Pr[n] = 1/N$, a result of the uniform pdf (1). The filtering gain $G(N)$ can be simply rewritten in the form

$$G(N) = 10 \log N - \frac{10}{N} \sum_{1}^{N} \log n. \tag{5}$$

Figure 7 plots $G(N)$ as a function of N.

The asymptotic value

$$G(\infty) = \int_0^{f_0} \left(10 \log \frac{f_0}{f_L(t)}\right) \cdot p(f_L(t)) d f_L(t) \tag{6}$$

can be evaluated simply by using the identity

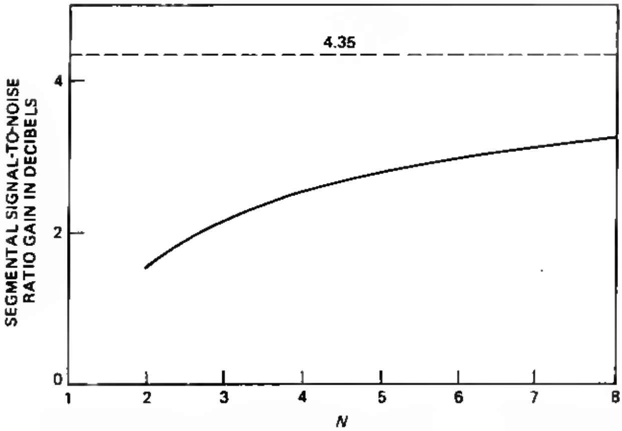

Fig. 7—Segmental s/n gain $G(N)$ versus N. This characteristic assumes uniform pdf models for coding noise and $f_L(t)$. The asymptotic value is $G(\infty) = 4.35$ dB. More than half this gain is realized with $N = 4$.

$$\int \ln z = z \ln z - z,$$

and this results in

$$G(\infty) = 10/e = 4.35 \text{ dB}. \tag{7}$$

This asymptotic value is indeed close to the ideal adaptive filtering gains reported in earlier-cited ADM experiments at several bit rates.[1] It is also seen from Fig. 6 that the four-filter bank method of Fig. 5c theoretically realizes more than one-half of the maximum possible dB gain $G(\infty)$ in segmental s/n: $G(4) \simeq 2.6$ dB.

The above analytical formulation can also be used to assess the efficiency of the equispaced filter design in (2) and Fig. 6. For illustration, $G(2) = 1.5$ dB with $N = 2$. It can be shown analytically that an optimal design, one that maximizes adaptive filtering gain for $N = 2$, is one for which $f_{c1} = (f_0/e)$, rather than $f_0/2$. Figure 8 plots the theoretical expected gain in segmental s/n as a function of f_{c1}. The maximum gain is 1.6 dB, only 0.1 dB better than that in the simple design of (2), which suggests $f_{c1} = f_0/2$ for $N = 2$.

IV. RESULTS OF COMPUTER SIMULATIONS

ADPCM coders with two- and four-bank filtering systems were simulated with the speech inputs mentioned in Section I. The coder bit rates were 24 and 16 kilobit/s, corresponding to 3- and 2-bit/sample quantizers, and an 8-kHz sampling of the inputs. The quantizers were adaptive. Both backward adaptive (AQB) and forward-adaptive (AQF) quantizers were simulated;[4] filtering gains were evident in both cases,

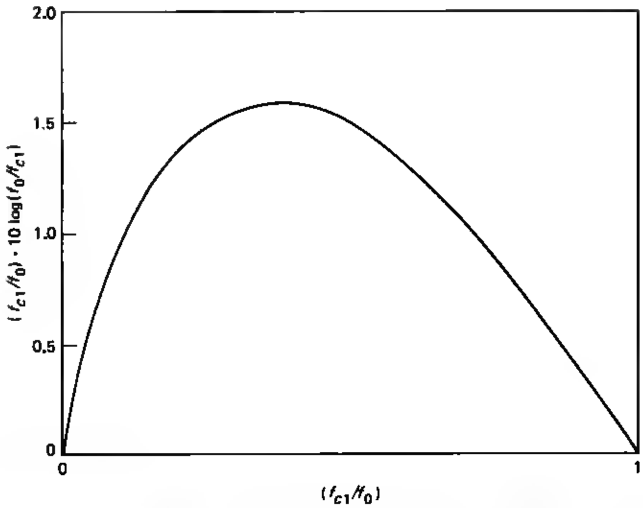

Fig. 8—Segmental s/n gain with $N = 2$ as a function of f_{CL}/f_0. The maximum value of 1.6 dB occurs for $f_{c1}/f_0 = e^{-1}$. The scheme of Fig. 6 suggests $f_{c1}/f_0 = 0.5$ for $N = 2$. This produces an s/n gain only 0.1 dB less than the maximum value of 1.6 dB.

but in terms of absolute quality, the AQF coders were clearly better, especially at 16 kilobit/s. AQF coders, however, require the explicit transmission of step-size information; using for example, four bits once for every segment of 256 samples (16 ms).

The cutoff frequency $f_L(t)$ was computed once for every segment of 256 samples. These segments were Hamming-windowed, zero-padded for better frequency resolution, and analyzed by means of a 512-sample FFT.

The filter banks consisted of 33-point FIR filters whose frequency responses are shown in Figure 9. Although gains due to a two-filter system were almost always noticeable, they were not always significant; specifically they were not significant for inputs with predominant all-pass ($>f_0/2$) segments; see Figs. 3 and 4. A four-filter bank, on the other hand, provided significant quality improvement with all the four test inputs. The perceived improvement in quality was much greater in all cases than what the theoretically predicted objective gain of $G(4) = 2.6$ dB indicates. The measured gains were very input-dependent with an average value slightly less than the theoretical value of $G(4)$. This result reinforces earlier results for time-varying filtering of ADM speech,[1] where again perceptual gains were in excess of objectively measured segment s/n improvements; this is probably because the residual in-band noise after adaptive filtering is much less annoying because of masking by tbe input signal. Adaptive bandwidth ADPCM, of course, has the additional possibility of variable bit allocation. For

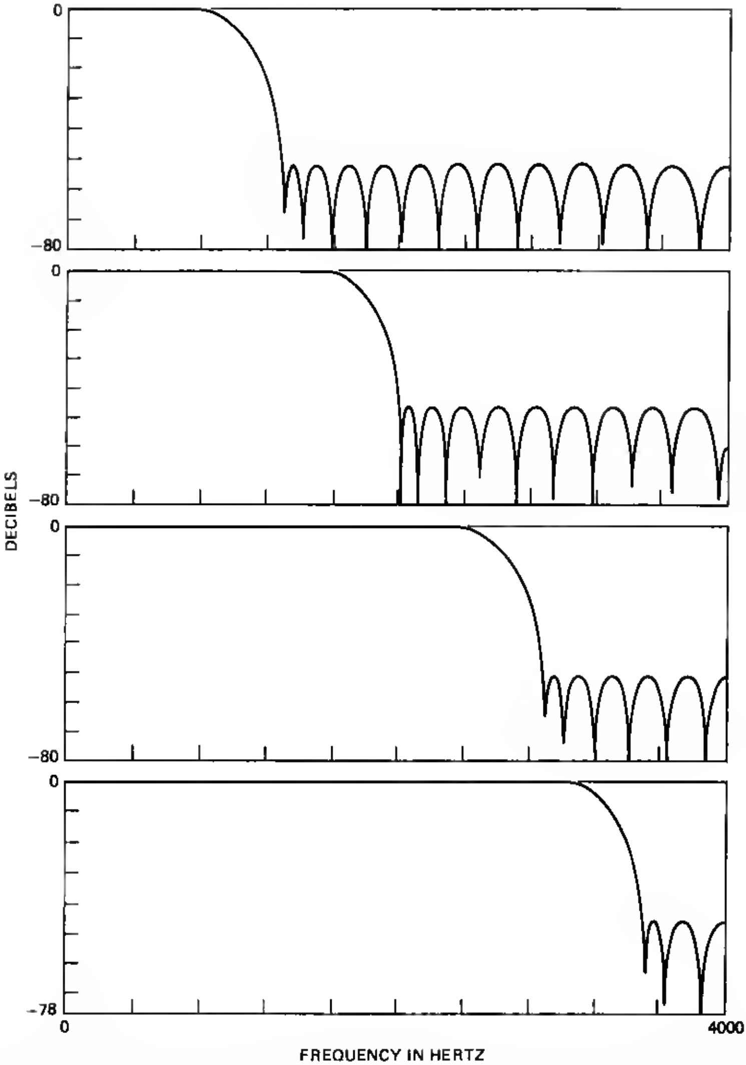

Fig. 9—Frequency responses of low-pass filters in an $N = 4$ filter bank. Each filter is a 33-point FIR design, and the cutoffs correspond to those in Fig. 5c and Fig. 6, with $f_0 = 3.2$ kHz.

example, significant quality improvements have been noticed in a system where the input was subsampled at 5.33 kHz whenever $f_L(t) < 2.4$ kHz, and more quantization bits were allocated for subsampled segments. For example, in 16 kilobit/s operation tbese segments were coded with three-bits/sample instead of two. This type of variable bit allocation is ruled out by definition in ADM which is a one-bit/

sample system. Notice finally that ADPCM with out-of-band noise rejection and variable bit allocation is very similar to frequency-domain subband coding.[2]

V. CONCLUSIONS

We have indicated a conceptually simple but very effective procedure for improving the quality of ADPCM speech coding. The extra information for effecting this improvement is very little; with a four-filter design this extra information would be $\log_2 4$ = two bits, corresponding to four possible ranges of cutoff frequency $f_L(t)$. The complexity involved in terms of computing $f_L(t)$ is quite significant in relation to the simplicity of the conventional, basic ADPCM coder. The increased complexity, however, will be relatively less objectionable in voice storage applications than in transmission systems; and in either case, an attractive feature is that the post-filtering procedure (the boxed portion of Fig. 6) can be used as an entirely optional refinement. Also, as noted earlier, the adaptive post-filtering technique discussed in this paper can be used with significant gains in the context of coders other than ADPCM.[1]

VI. ACKNOWLEDGMENT

This work was prompted by a summer project on post-filtered ADM by J. B. Allen and J. O. Smith.[1]

REFERENCES

1. J. O. Smith and J. B. Allen, "Variable Bandwidth Adaptive Delta Modulation," B.S.T.J., this issue.
2. J. L. Flanagan et al., "Speech Coding," IEEE Trans. Commun., *COM-27*, No. 4 (April 1979), pp. 710–37.
3. N. S. Jayant, "A First-Order Markov Model for Delta-Modulation Noise Spectra," IEEE Trans. Commun. (August 1978), pp. 1316–18.
4. N. S. Jayant, "Step-Size Transmitting Differential Coders for Mobile Telephony," B.S.T.J., *54*, No. 9 (November 1975), pp. 1557–81.